

# School-Based Mental Health and Behavioral Programs for Low-Income, Urban Youth: A Systematic and Meta-Analytic Review

Farahnaz K. Farahmand, Kathryn E. Grant, Antonio J. Polo, and Sophia N. Duffy, DePaul University

David L. DuBois, University of Illinois

**A systematic and meta-analytic review was conducted of the effectiveness of school-based mental health and behavioral programs for low-income, urban youth. Applying criteria from an earlier systematic review (Rones & Hoagwood, 2000) of such programs for all populations indicated substantially fewer effective programs for low-income, urban youth. The meta-analysis similarly failed to indicate effects of the typical program on primary outcomes. Effectiveness was evident, however, for programs that targeted internalizing problems or had a broader socio-emotional focus and those delivered to all youth (i.e., universal). In contrast, negative effects were apparent for programs that targeted externalizing problems and were delivered selectively to youth with existing problems. Distinctive characteristics of low-income, urban schools and nonschool environments are emphasized as potential explanations for the findings.**

**Key words:** interventions, low-income, mental health services, schools, urban, youth. [*Clin Psychol Sci Prac* 18: 372–390, 2011]

More than 39.8 million people live in poverty in the United States. Of these, more than 14 million are youth under the age of 18 (U.S. Bureau of the Census,

2011). Youth of color are especially likely to be poor. Poverty rates for African American youth are almost two and a half times those for European American youth and poverty rates for Latino youth are nearly two times those for European American youth, with approximately 34.4% of African American and 30.3% of Latino youth living in poverty (U.S. Bureau of the Census, 2011). Youth of color are also more likely to live in segregated urban communities where there are few resources and high rates of unemployment, homelessness, and crime (U.S. Department of Health and Human Services, 2001).

Poverty is associated with stressors that range from major life events (e.g., child abuse, divorce) to chronic interpersonal stressors (e.g., family conflict) to daily hassles (e.g., lack of money for transportation; Conger, Ge, Elder, Lorenz, & Simons, 1994). Additionally, living in urban poverty brings increased exposure to crime and violence, particularly for adolescents (e.g., Bell & Jenkins, 1994). For example, more than 40% of inner-city youth have seen someone shot or stabbed (U.S. Department of Health and Human Services, 2001). Traumatic and stressful experiences, in turn, have been established as risk factors for a range of psychological problems (e.g., Grant et al., 2003).

It is not surprising, therefore, that low-income, urban youth are at heightened risk for a number of psychological problems (Grant et al., 2004). For example, based on the normative data for the Youth Self-Report (YSR; Achenbach, 1991), approximately 5% of Grant and

Address correspondence to Farahnaz K. Farahmand, Department of Psychology, DePaul University, 2219 N. Kenmore Ave., Chicago, IL 60614. E-mail: ffarahma@depaul.edu.

colleagues' (2004) predominantly minority sample of low-income, urban youth were expected to score in the clinical range. However, significantly more adolescents scored in the clinical range on each of the YSR syndrome subscales, including Internalizing (24% of girls; 28% of boys), Externalizing (35% of girls; 26% of boys), Anxious-Depressed (7% of girls; 9% of boys), and Delinquent Behavior (17% of girls; 16% of boys).

Although low-income, urban youth are at higher risk for the development of psychological problems, they are less likely to receive help (Farmer, Stangl, Burns, Costello, & Angold, 1999; Garland et al., 2005). A recent study found that nearly 80% of low-income youth in need of mental health services had not received services within the preceding 12 months (Kataoka, Zhang, & Wells, 2002). Additionally, those who do receive mental health services experience attrition rates >50% because of a number of practical and structural barriers such as stigma, lack of information, inaccessible location of services, or difficulty with transportation (Kazdin, Holland, & Crowley, 1997). Given these barriers and the lack of services provided to youth most at need, the education sector has played a central role as an entry point into the mental health system for low-income, urban youth (Farmer, Burns, Phillips, Angold, & Costello, 2003; U.S. Department of Health and Human Services, 1999).

Comprehensive health services, including services for mental health, were first offered in schools in the mid-1980s (Dolan, 1992). Since then, the Surgeon General's Report (U.S. Department of Health and Human Services, 1999) has described schools as a key setting for the identification and treatment of mental disorders in children and youth. Data from the 2004–2005 national survey of 1,235 school-based health centers (SBHCs) revealed that 65% of SBHCs have services provided by mental health staff and 59% of the centers are located within urban schools (National Assembly on School-Based Health Care, 2007). Nonetheless, the type or quality of services being offered in school settings is poorly understood (Rones & Hoagwood, 2000). Extant research suggests school-based mental health services are being delivered in a variety of ways without an explicit "best practice" model (Paternite, 2005).

A systematic review of published reports of school-based mental health services spanning 1985–1999 was com-

pleted by Rones and Hoagwood (2000). These authors defined school-based mental health services as "any program, intervention, or strategy applied in a school setting that was specifically designed to influence students' emotional, behavioral, or social functioning." The results of that review were organized around the target of the intervention (e.g., depression, conduct problems) and within each of these categories, the level at which the program/service intervened. Both universal (administered to all youth) and selective/indicated programs (administered to youth with identified problems) were included; thus, both treatment and prevention studies were examined.

Forty-seven studies were included in the Rones and Hoagwood (2000) review. Overall, 17 (36%) were classified as effective, 17 (36%) as mixed, and nine as ineffective (28%). More specifically, five studies focused on emotional and behavioral problems more generally, with three identified as being effective and two with mixed results; six studies focused on depression, with three effective, one mixed, and two listed as not effective; 22 studies focused on conduct problems, with eight identified as effective, 10 with mixed results, and four not effective; two studies, both effective, focused on stress management; finally, 12 studies focused on substance use, with three effective, six mixed, and three not effective. These results suggest that there are a number of school-based mental health programs that have evidence of impact across a range of emotional and behavioral problems. Rones and Hoagwood (2000) identified five key characteristics of school-based programs that positively affected outcomes, service sustainability, and maintenance, including (a) consistent program implementation; (b) inclusion of parents, teachers, or peers; (c) use of multiple modalities; (d) integration of program content into general classroom curriculum; and (e) developmentally appropriate program components.

The purpose of the current review is to build upon the Rones and Hoagwood (2000) review by systematically examining school-based mental health services and programs implemented with low-income, urban youth, given (a) this population is at particular risk for developing psychological problems as a result of the chronic and severe stressors they face, (b) they have limited access to services outside a school setting, and

(c) published reviews in this area have not focused on the socioeconomic characteristics of study samples and no review to date has specifically focused on low-income youth living in urban settings.

We chose to replicate the Rones and Hoagwood (2000) approach and reviewed interventions that targeted a spectrum of disorders for several reasons: (a) this allowed us to compare our findings with theirs; (b) low-income, urban youth have been shown to be at heightened risk for a range of psychological problems (Grant et al., 2004); (c) a number of school-based interventions target multiple outcomes (see Rones & Hoagwood, 2000, and this review); and (d) there is a growing trend in intervention research to develop interventions that simultaneously target a range of psychological problems (McHugh, Murray, & Barlow, 2009). The current review included all studies previously identified in the Rones and Hoagwood (2000) review that were conducted with low-income, urban youth, and updated the literature with studies published from 2000 through 2009 that met their specific inclusion criteria.

The current review addresses two primary questions in an effort to inform best practices for mental health services and program implementation for low-income, urban youth within school settings: (a) How effective have school-based mental health and behavioral programs been in promoting positive outcomes for low-income, urban youth? More specifically, to what extent have these programs had a favorable impact on outcomes for low-income, urban youth and to what extent are impacts sustained over time? (b) What factors influence the effectiveness of school-based mental health and behavioral programs for low-income, urban youth? More specifically, do program effects vary based on the type of problem that is targeted (e.g., conduct disorder vs. depression), the characteristics of the intervention (e.g., universal vs. selected), the characteristics of the sample (e.g., ethnic composition), or the factors that Rones and Hoagwood (2000) concluded were important based on their narrative review of the literature on school-based mental health and behavioral programs for general populations of youth?

To apply the most *rigorous* approach to answering these questions, we conducted a meta-analysis of all studies that met our inclusion criteria and provided data necessary to calculate effect sizes (see Table 1). To

apply the most *comprehensive* approach to answering these questions and to allow for a more direct comparison with results of the Rones and Hoagwood (2000) review, we also qualitatively categorized the effectiveness of studies that met our inclusion criteria but did not provide data necessary to calculate effect sizes.

## METHOD

### Inclusion Criteria

To be included in our review, the program evaluated had to meet Rones and Hoagwood's (2000) definition of "school-based mental health services" provided earlier in this article. Consistent with this definition, programs designed to influence only academic achievement were excluded; however, academic-related outcomes when assessed by eligible studies were included (e.g., Walker, Kerns, Lyon, Bruns, & Cosgrove, 2010). We excluded mentoring programs based in schools, as no interventions of this type were included in the Rones and Hoagwood review and because the effectiveness of such programs has been reviewed elsewhere (e.g., Wheeler, Keller, & DuBois, 2010). Consistent with Rones and Hoagwood, we included school-based studies with additional nonschool components (e.g., family intervention) and specifically tested whether these added components were associated with stronger effects.

We also used the same methodological inclusion criteria used by Rones and Hoagwood (2000). Specifically, only those studies that included a control group, standardized outcome measures, and outcomes assessed at baseline and postintervention were eligible. Three designs were considered acceptable: (a) randomized designs; (b) quasi-experimental designs that used multiple sites and demographically matched samples to minimize selection biases; and (c) multiple baseline designs using sample cohorts as their own controls (Rones & Hoagwood, 2000). Although the inclusion criteria for control groups were not specified by Rones and Hoagwood, all studies with groups intended to serve as controls were included in this review.

Additional inclusion criteria were as follows. First, studies had to be conducted with samples in the United States to ensure uniformity in the definition of low income and to avoid confounding of findings with cross-cultural differences in school contexts.

Table 1. School-based mental health programs for low-income students with effect size calculations

| Study   | Program Title                         | Total N (Treatment Group n)  | Grade                                  | Sex (% Female)    | Ethnicity  | Target Problem                            | Delivery Agent | Mean Effect Size (g)—Primary Outcomes*   |
|---|---------------------------------------|--|--|-------------------|--|---|----------------|--|
| Aber et al. (1998)  | Resolving Conflict Creatively Program | 5,053 (not specified)  | 2nd–6th                                | 48%               | 43% Latino; 36% African American; 16% European American; and 5% Not specified                                    | Conduct Problems (U)                      | UC—teachers    | Post: 0.05   |
| Abt Associates (2001)                                     | Families and Schools Together         | 382 (193)  | 2nd–4th                                | 38%               | 90% African American and 10% Not specified   | Conduct Problems (S)                      | UC—teachers    | Post: 0.05   |
| Botvin et al. (2001)                                      | Life Skills Training                  | 3,621 (2,144)  | 7th–8th                                | 53%               | 61% African American; 22% Latino; 6% Asian American; 6% European American; and 5% Mixed/Not specified            | Substance Use (U)                         | UC—teachers    | Post: 0.07<br>F/U: 0.10  |
| Cardemil et al. (2002, 2007)                              | Penn Resiliency Program               | I: 49 (23);<br>II: 103 (47)  | 5th–6th                                | I: 45%<br>II: 56% | I: 100% Latino<br>II: 100% African American  | Depression (U)                            | R              | Latino<br>Post: 0.93<br>F/U1: 0.63<br>F/U2: 0.97<br>African American<br>Post: 0.16   |
| Cho et al. (2005)   | Re-connecting Youth                   | 1,218 (617)  | 9th–11th                               | 50%               | 47% Latino; 24% Asian American; 13% African American; 9% European American; and 7% American Indian/Not specified | Conduct Problems (S)                      | UC—teachers    | Post: $\downarrow$ 0.02<br>F/U: $\downarrow$ 0.10  |
| Henderson et al. (1992)                                   | Stress-Control Program                | 65 (33)  | 3rd                                    | 57%               | 60% African American and 40% European American   | Broad Mental Health and/or Behavioral (U) | R              | Post: 0.76   |
| Hostetler and Fisher (1997)                               | Project C.A.R.E.                      | 317 (163);<br>F/U1: 187 (97)<br>F/U2: 78 (39)  | 4th                                    | 43%               | 39% Latino; 34% European American; and 27% African American  | Substance Use (S)                         | R              | Post: $\downarrow$ 0.16<br>F/U1: 0.16<br>F/U2: 0.03  |
| Jagers et al. (2007)                                      | Aban Aya Youth Project                | 789 (417)  | 5th–7th                                | 51%               | Predominately African American (% unknown)   | Broad Mental Health and/or Behavioral (U) | R              | Group 1<br>Post: 0.19<br>Group 2<br>Post: 0.18   |
| McClowry et al. (2005)                                    | INSIGHTS into Children's Temperament  | 148 (91)   | 1st–2nd                                | 34%               | 89% African American; 9% Latino; and 2% Mixed/Not specified  | Broad Mental Health and/or Behavioral (U) | R              | Post: 0.61   |
| McDonald et al. (2006)                                    | Families and Schools Together         | 130 (80)   | 1st–4th                                | 73%               | 100% Latino  | Broad Mental Health and/or Behavioral (U) | UC—teachers    | Post: 0.63   |
| Metropolitan Area Child Study Research Group (MACS, 2002) | MACS intervention                     | Early Levels:<br>A: 152 (120)<br>B: 171 (82)<br>C: 238 (120)<br>Late Levels:<br>A: 103 (52)<br>B: 67 (37)<br>C: 100 (49) | Early:<br>2nd–3rd;<br>Late:<br>5th–6th | 39%               | 48% African American; 37% Latino; and 15% European American  | Conduct Problems (S)                      | UC—teachers    | Early<br>Post: A: 0.11<br>Post: B: $\downarrow$ 0.17<br>Post: C: $\downarrow$ 0.41<br>Late<br>Post: A: $\downarrow$ 0.36<br>Post: B: $\downarrow$ 0.50<br>Post: C: $\downarrow$ 0.27 |

(Continued)

Table 1. (Continued)

| Study                      | Program Title   | Total N (Treatment Group n) | Grade                  | Sex (% Female) | Ethnicity  | Target Problem                            | Delivery Agent  | Mean Effect Size (g)—Primary Outcomes* |
|----------------------------|---|-----------------------------|------------------------|----------------|--|---|---|--|
| Mufson et al. (2004)       | Interpersonal Psychotherapy for Depressed Adolescents | 63 (34)                     | Middle and High School | 84%            | 71% Latino and 29% Not specified                                   | Internalizing Symptoms (S)                | UC—school-based health clinic therapists                        | Post: 0.42                             |
| Murray and Malmgren (2005) | N/A   | 48 (24)                     | 9th–12th               | 23%            | 100% African American  | Broad Mental Health and/or Behavioral (S) | UC—teachers   | Post: 0.25                             |
| Sinclair et al. (2005)     | Check and Connect                                     | 144 (71)                    | High School            | 18%            | 67% African American; 22% European American; and 11% Not specified | Broad Mental Health and/or Behavioral (S) | R   | Post: 0.38                             |
| Tolan et al. (2004)        | SAFEChildren  | 442 (243)                   | 1st                    | 49%            | 58% Latino and 42% African American                                | Conduct Problems (U)                      | R   | Post: ; 0.05<br>F/U: ; 0.06            |
| Weiss et al. (2003)        | Reaching Educators, Children & Parents                | 93 (62)                     | 4th                    | 37%            | 56% African American; 38% European American; and 6% Not specified  | Broad Mental Health and/or Behavioral (S) | R: outside master's-level social workers and psychiatric nurses | Post: 0.11<br>F/U: 0.28                |

Note. Post = Post-treatment; F/U1 = First follow-up; F/U2 = Second follow-up; U = Universal; S = Selected; UC = Usual care provider; R = Research staff.

\*Effect sizes were computed so that positive values indicated differences in directions consistent with a favorable effect of the intervention group on youth outcomes (e.g., higher self-esteem, fewer symptoms of depression).

Second, the study sample had to be predominately or exclusively low income and urban. A study was judged to meet the urban portion of this criterion if it stated its sample was drawn from schools in an urban area or from a city deemed to be an urban area (U.S. Bureau of the Census, 2000). If a sample included a mixed urban and suburban/rural population, it was excluded unless results were clearly parsed out for the urban youth and these youth met the poverty inclusion criteria. The poverty criterion was met if more than 60% of the sample was described as low income based on measures such as eligibility for free school lunch programs or economic indexes. Additionally, when this information was not reported, studies were included if terms such as "all," "majority," or "most" were used to describe the sample's level of poverty or if urban poverty was discussed in the context of an "inner-city" sample, which, as defined, includes both urban and impoverished elements. However, if the city was described as impoverished but the study sample was described as <60% impoverished, the study was excluded. Third, as was the case with the Rones and Hoagwood (2000) review, studies had to be published in a peer-reviewed journal; accordingly, theses and dissertations were not included. Fourth, studies needed to either be included in the Rones and Hoagwood review or published in the period of 2000–2009.

#### Literature Search

Although only one study was explicitly described as implemented with a low-income, urban sample in text, our examination of the 47 articles reviewed by Rones and Hoagwood (2000) revealed that eight of the evaluations were conducted with a sample of predominately low-income, urban youth. These studies were included in the current review. In addition, a computerized search of references published between 2000 and 2009 identified from PsycINFO and a manual search (tracking citations within identified articles) were used to update the literature search, resulting in a review that spans 1985–2009. Search terms used by Rones and Hoagwood included "schools, children, mental health, services, prevention, outcomes, effectiveness, and specific syndromes (e.g., ADHD, depression), among others." Key words for identifying studies between

2000 and 2009 included those terms in addition to the following: psychopathology (or outcomes), youth (or adolescents), urban (or inner-city), low-income (or impoverished or poverty), and services (or intervention or risk reduction). Various combinations of these terms were used to cast the broadest net possible.

These search procedures resulted in a final sample of 23 articles evaluating 19 programs across 21 studies and 29 independent samples. Most studies included both men and women, with two studies (Murray & Malmgren, 2005; Sinclair, Christenson, & Thurlow, 2005) including a predominately male (over 75%) sample. A total of 20 studies (95% CI) used random assignment in their design (12 individual level; one classroom level; seven school level). The remaining study used a quasi-experimental design. Study designs included intervention conditions compared with no treatment controls (86%), attention/placebo controls (10%), and waitlist controls (4%).

#### Study Coding

Two clinical psychology doctoral students coded the studies. Both were trained and supervised by doctoral-level psychologists, one of whom completed independent coding of selected variables (e.g., target focus of the program), with discrepancies from original coding resolved by consensus.

*Target Focus.* Programs were coded as primarily focused on conduct, depression, substance use, or broad mental health and/or behavioral targets. Studies pulled from Rones and Hoagwood's (2000) review maintained the same target focus code as applied in that review with the exception of one study, which was categorized as "stress management" in the Rones and Hoagwood review but categorized as broad mental health and/or behavioral in this review (Henderson, Kelbey, & Engerbreston, 1992).

*Level of Intervention.* Studies were categorized as universal if delivered to all youth; selected if delivered to youth selected on the basis of symptoms, behaviors, or other criteria short of diagnosis; and indicated if provided to youth who met diagnostic criteria for a mental health disorder (Institute of Medicine, 1994). As there were fewer selected and indicated interventions, these categories were combined in a selected/indicated category.

This categorization is consistent with the one used in the Rones and Hoagwood (2000) review.

*Change Agent.* The change agent refers to the person who delivered the program. Change agent categories were researchers/research-hired or usual care providers (e.g., teachers).

*Ethnicity.* Studies were categorized based on the predominant (>75%) ethnic composition of their sample. Codes included predominantly African American and predominantly Latino. Although Asian American, American Indian, and European American students were enrolled in a number of eligible studies, no study had 75% or greater enrollment of any of these groups. Multiple minority was coded if at least 75% of the sample was non-European American but no single group made up 75% of the sample. Multiple ethnicity was coded if European American youth were included in the sample and the combined ethnic minority enrollment was below 75% of the sample. (Note: There were two exceptions to this categorization. One study comprised 71% Latino youth, with no other information on ethnicity included in the predominantly Latino group, and one study with a sample of 72% minority youth without specification of ethnic groups was included in the multiple minority category.)

*Consistent Program Implementation.* Implementation can refer to (a) fidelity, or the extent to which a program is delivered as it was intended; (b) dosage, or the number of sessions delivered or completed; or (c) quality, or how well different program components are delivered (Durlak & DuPre, 2008). Studies were coded on whether or not implementation was assessed, and, if so, the reported quality of the study's implementation (low, adequate, high).

*Inclusion of Parents, Teachers, and Peers.* Programs were coded to indicate whether or not parents, teachers/other school staff, and/or peers were incorporated.

*Multiple Modalities/Components.* Programs were further coded as to whether they included components beyond a school-based curriculum (e.g., also a family intervention).

*Integration Into Classroom.* Integration into the classroom was coded based on the definition provided by Rones and Hoagwood (2000): "delivered as an integral part of the classroom curricula rather than as a separate and specialized lesson."

*Developmental Timing.* Rones and Hoagwood (2000) coded the concepts and curricula based on whether they were developmentally appropriate. In the present review, an analysis of available materials suggested this task would be too speculative, so it was dropped. Instead, the current review examined developmental timing of programs by using age as a moderator. Average age of participants was examined both continuously and as a categorized grade variable: elementary (ages 5–10), middle school (ages 11–13), and high school (ages 14–18).

*Outcomes.* Primary outcomes were those related to the target of the program. Internalizing outcomes were deemed primary for depression-focused programs; externalizing outcomes were deemed primary for conduct-focused programs; substance use outcomes were deemed primary for substance use programs; and externalizing, internalizing, competence/social skills, and academic outcomes were deemed primary for broad mental health and/or behavioral-focused programs (see Table 1). All other outcomes were coded as secondary.

#### Data Analytic Procedures

*Meta-Analytic Procedures.* Effect sizes were computed as standardized mean differences. This involves taking the raw difference between treatment and control group means on the outcome measure at post-treatment and then dividing this difference by the pooled (weighted average) standard deviation of the measure for the two groups (see Cooper, 2010, formula 5.11). An adjustment developed by Hedges (1981) was also incorporated to address bias that can arise in calculation of this type of effect size with small samples (Hedges's *g*). Pretest means were subtracted from post-treatment means to adjust for potential differences between program and comparison groups at baseline and thus enhance precision in effect size estimation (Lipsey & Wilson, 2001). In most instances, effect sizes were computed from means and standard deviations on outcome measures that were

included in the study report or made available to us by study authors. When these were not available, effect sizes were estimated from relevant test statistics or their reported significance levels (see Rosenthal, 1994). Effect sizes were computed so that positive values indicated differences in directions consistent with a favorable effect of the intervention group on youth outcomes (e.g., higher self-esteem, fewer symptoms of depression).

The independent sample was the primary unit of analysis. Because effect size information was reported for the overall sample in most reports, each report or study generally contributed one independent sample to the analysis. In some instances, however, studies only reported findings separately for nonoverlapping subgroups (e.g., Latino vs. African American youth; classroom only vs. family also conditions). These studies contributed more than one sample to the analysis (i.e., one sample for each distinct subgroup; Cooper, 2010). Furthermore, some studies included two or more distinct interventions compared against a control. In these instances, we computed effect sizes for each intervention-control group comparison and treated the corresponding samples as independent in study analyses. Finally, we also allowed studies to contribute information to each outcome category for which effect size information was available.

Effect sizes that were more than three interquartile ranges above the 75th percentile or below the 25th percentile, and thus qualified as statistical outliers according to Tukey's definition (Tukey, 1977), were Winsorized by setting their values to the highest or lowest effect size, respectively, that did not qualify as an outlier. Doing so provided a safeguard against extreme effect sizes having undue influence on our findings. In addition, each effect size was weighted by the inverse of its variance to provide more efficient estimation of true population effects (Hedges & Olkin, 1985). This procedure gives greater weight to larger samples and is the generally preferred approach (Cooper, 2010). When evaluations used data that were clustered across multiple units (e.g., schools), effect size weights were adjusted to account for such clustering using procedures recommended by Higgins and Green (2008). This adjustment makes use of the intra-class correlation (ICC) coefficient for the outcome measure. Because ICC values were generally not reported in

studies, we substituted our own estimate of 0.10 (What Works Clearinghouse, 2008).

Analyses were conducted with a random-effects model (Hedges & Vevea, 1998). When a random-effects analysis is carried out, a study-level variance component is assumed to be present as an additional source of random influence on effect sizes. In general, a random-effects analysis is more conservative because of the consideration of study-level variance as an additional component of error (Lipsey & Wilson, 2001). The appropriateness of a random-effects model for the present analysis was indicated by (a) substantial variability in the characteristics and participants of youth in the included programs and the potential for such differences to constitute significant sources of random error even after taking into account variance associated with specified moderating variables and (b) interest in drawing inferences about all programs, not just those included in the present review (Hedges & Vevea, 1998).

To test whether there was variability in sample-level effect sizes greater than that which would be expected by sampling error around a single population value, we conducted a homogeneity analysis using procedures described by Cooper (2010). Results of this analysis were used as well to calculate  $I^2$ , a descriptive measure of the amount of the observed variability in effect sizes across studies that is attributable to study differences rather than sampling error (Higgins & Thompson, 2002). The homogeneity analysis and all moderator analyses were completed with post-treatment effect sizes only, given the limited number of studies that were assessed at follow-up time points. Because of the potential for confounding among study characteristics tested as moderators in a meta-analysis (Cooper, 2010), for any moderator that demonstrated a significant association with effect size we tested whether the association remained evident when controlling, in turn, for each of the other study characteristics tested as moderators.

*Qualitative Procedures.* Criteria used to label studies as effective, mixed, and ineffective were not clearly defined by Rones and Hoagwood (2000). But, it appears effective studies reported only significant positive outcomes and studies listed as ineffective reported only non-significant and/or negative treatment effects. Detailed results were not provided for effective and ineffective



studies, but they were provided for studies with mixed results. In this mixed category, results were significant and positive for some but not other outcomes.

In the present review, significant positive, nonsignificant, and significant negative treatment effects were used to classify all studies (including those that did not provide data necessary for calculating effect sizes) as effective, ineffective, or mixed for a more direct comparison with Rones and Hoagwood's findings. Studies were categorized as effective if they reported >75% positive post-treatment outcomes related to the program's focus, mixed if between 50% and 74% of program-focused post-treatment outcomes were positive, and ineffective if <50% of program-focused post-treatment outcomes were positive.

## RESULTS

Results are presented by research question. The first question pertaining to overall program effectiveness was addressed using both qualitative and meta-analytic procedures, whereas the second question was addressed using meta-analysis only. Table 1 lists the interventions reviewed meta-analytically and provides study information, including sample characteristics and results. Effect sizes are listed for the overall sample.<sup>1</sup>

**How Effective Have School-Based Mental Health and Behavioral Programs Been in Promoting Positive Outcomes for Low-Income, Urban Youth? More Specifically, to What Extent Have These Programs Had a Favorable Impact on Outcomes for Low-Income, Urban Youth and to What Extent Are Impacts Sustained Over Time?**

*Qualitative Analyses.* Qualitative analyses of the 29 samples included in this review<sup>2</sup> resulted in five programs classified as effective (17%), eight as mixed (28%), and 16 as ineffective (55%). Of the conduct-focused programs, no programs were deemed effective, three were deemed mixed, and nine were deemed ineffective. Of the depression-focused programs, one was deemed effective, one mixed, and one ineffective. Of the substance use-focused programs, one was deemed effective and three ineffective with no mixed programs. Finally, of the general mental health and behavioral-focused programs, three were deemed effective, four mixed, and three ineffective. For the universal programs, four were effective, four mixed, and six

ineffective. For the selected programs, one was effective, five mixed, and 10 ineffective.

*Meta-Analysis.* Of the 29 independent samples, 23 had at least one calculated effect size at post-treatment, six at a first follow-up time point (average length of time 10 months), and two at a second follow-up time point (average length of time 12 months). No effect sizes were able to be calculated beyond the second follow-up time point. The overall aggregated effect sizes for the primary outcomes were 0.08 (95% CI ) 0.01 to 0.17) at post-treatment, 0.06 (95% CI ) 0.07 to 0.20) at first follow-up, and 0.48 (95% CI ) 0.44 to 1.41) at second follow-up. The mean difference between effect sizes at post-treatment and first follow-up for the same outcome measures was 0.03 (95% CI ) 0.08 to 0.14), indicating maintenance of gains over time for the intervention group. The mean difference between post-treatment and second follow-up for the same outcome measures was 0.17 (95% CI ) 0.04 to 0.37), indicating additional gains over time for the intervention group; however, because of small sample size at this time point, these gains should be interpreted with caution.

The difference at post-treatment between the aggregated effect sizes across the primary outcome categories was significant ( $Q_b = 10.79$ ,  $df = 3$ ,  $p = .01$ ). Given the low number of studies with substance use outcomes, these outcomes were combined with the externalizing outcomes for all subsequent analyses. At post-treatment, outcomes assessing competence and social skills had the highest mean effect size (Mean ES = 0.31, 95% CI 0.13–0.50,  $k$  [number of independent samples] = 5), followed by internalizing outcomes (Mean ES = 0.28, 95% CI 0.05–0.51,  $k = 6$ ), academic outcomes (Mean ES = 0.24, 95% CI ) 0.02 to 0.50,  $k = 4$ ), and finally externalizing/substance use outcomes (Mean ES = 0.02, 95% CI ) 0.08 to 0.12,  $k = 19$ ). Differences at follow-up time points could not be calculated as there were too few studies for this analysis.

Analyses with secondary outcomes were also conducted. The overall aggregated effect sizes for the secondary outcomes were 0.17 (95% CI 0.02–0.32,  $k = 14$ ) at post-treatment, 0.02 (95% CI ) 0.07 to 0.11,  $k = 6$ ) at first follow-up, and 0.02 (95% CI ) 0.25 to 0.30,  $k = 3$ ) at second follow-up. The mean difference between effect sizes at post and first

follow-up for the same secondary outcome measures was 0.00 (95% CI ) 0.06 to 0.06) and the mean difference between post-treatment and second follow-up for the same secondary outcome measures was 0.02 (95% CI ) 0.17 to 0.21), indicating neither deterioration nor gain in effects over time for the intervention group relative to the control group.

Analyses comparing mean effect sizes across secondary outcome categories were also conducted. Studies with secondary externalizing outcomes ( $n = 2$ ) and studies with secondary substance use outcomes ( $k = 2$ ) were combined to increase sample size, although still resulting in a lower than needed sample. Studies with internalizing outcomes ( $k = 3$ ) and studies with competence/social skills outcomes ( $k = 6$ ) were also combined to increase sample size. Academic studies had a sufficient number of samples to examine on their own as secondary outcomes ( $k = 9$ ). A marginally significant difference was found for these outcome groups ( $Q_b = 4.96$ ,  $df = 2$ ,  $p = .08$ ), with secondary externalizing/substance outcomes having the lowest and negative effect size (Mean ES = ) 0.05, 95% CI ) 0.25 to 0.16), secondary internalizing/competence outcomes having the next highest yet still small effect size (Mean ES = 0.08; 95% CI ) 0.09 to 0.24), and academic outcomes having the highest, yet still small effect size (Mean ES = 0.24, 95% CI 0.08–0.40).

#### **What Factors Influence the Effectiveness of School-Based Mental Health and Behavioral Programs for Low-Income, Urban Youth?**

The homogeneity analysis revealed significant variation in effect sizes across samples,  $Q = 74.63$ ,  $k = 23$ ,  $p < .001$ . The corresponding  $I^2$  value of 70.5% reflects a medium to large degree of heterogeneity (Higgins & Thompson, 2002). These findings provide an empirical rationale for testing potential moderators that might account for this variation (Cooper, 2010). In preliminary analyses of study methodological characteristics as potential moderators, it was determined that there was a nonsignificant trend ( $Q_b = 1.97$ ,  $df = 1$ ,  $p = .16$ ) for effect sizes to be larger for samples in which youth were assigned to condition at the individual level (Mean ES = 0.18, 95% CI 0.01–0.35,  $k = 11$ ) rather than at the classroom or school level (Mean ES = 0.01, 95% CI ) 0.14 to 0.17,  $k = 12$ ). All subsequent analyses thus controlled for this methodological characteristic.

**Target Problem.** For the test of target problem as moderator, internalizing-focused interventions (Mean ES = 0.31, 95% CI 0.02–0.61,  $k = 3$ ) were combined with those with broad mental health and/or behavioral targets (Mean ES = 0.31, 95% CI 0.19–0.44,  $k = 8$ ), and conduct-focused interventions (Mean ES = ) 0.11, 95% CI ) 0.22 to ) 0.01,  $k = 10$ ) were combined with programs targeting substance use (Mean ES = 0.01, 95% CI ) 0.16 to 0.17,  $k = 2$ ). Results indicated a significant difference in effect size as a function of target problem ( $Q_b = 25.27$ ,  $df = 1$ ,  $p < .001$ ), with a positive and significant effect size when the program had an internalizing or broad/socio-emotional focus (Mean ES = 0.32, 95% CI 0.19–0.44,  $k = 11$ ) and a negative and nonsignificant effect size when the program had a focus on conduct problems or substance use (Mean ES = ) 0.09, 95% CI ) 0.18 to 0.01,  $k = 12$ ). This difference remained significant in all analyses controlling for other moderators.

**Level of Intervention.** Ten samples received universal and 13 received selected levels of intervention. Meta-analytic results revealed a significant difference in effect associated with this characteristic ( $Q_b = 12.67$ ,  $df = 1$ ,  $p = .002$ ), with interventions at the selected level (Mean ES = ) 0.08, 95% CI ) 0.21 to 0.04,  $k = 13$ ) having a smaller, nonsignificant effect size relative to those at the universal level for which there was a significant positive effect (Mean ES = 0.25, 95% CI 0.11–0.38,  $k = 10$ ). This difference remained significant in all analyses controlling for other moderators.

**Change Agent.** Thirteen samples received programs delivered by usual care providers and 10 received programs delivered by researchers or research-hired staff. Of those delivered by usual care providers, 44% of the samples received programs delivered by teachers. The remaining programs were delivered by other school staff such as counselors or a combination of providers. Initially, meta-analytic results revealed a significant difference associated with this characteristic; however, this difference was no longer significant when controlling for either the target problem addressed or the level of intervention. Use of researchers or research-hired staff to deliver the intervention was correlated strongly with programs having an inter-

nalizing or broad socio-emotional focus ( $r = 0.76$ ) and with programs being universal rather than selected ( $r = 0.31$ ).

**Ethnic Composition.** Of the samples included in the meta-analysis, six were primarily African American, three were primarily Latino, 10 were multiple minority, and four were multiple ethnicity. Although effect sizes for samples including predominantly Latino youth (Mean ES = 0.49) were more than twice as large as those that included predominantly African American youth (Mean ES = 0.23), there were insufficient samples in these categories to examine them using moderation analyses. Therefore, samples in which youth were predominantly African American and samples in which youth were predominantly Latino were combined and compared with a group made up of samples in which one type of race/ethnicity did not predominate. There was not a significant difference in effect size when race/ethnicity was examined as a moderator in this way ( $Q_b = 0.06$ ,  $df = 1$ ,  $p = .81$ ).

**Consistent Program Implementation.** Twelve of the 23 evaluations were assessed for various aspects of the implementation process. Analysis of possible differences in effects as a function of whether assessment of program implementation was reported (regardless of level) revealed a significant difference ( $Q_b = 5.30$ ,  $df = 1$ ,  $p = .02$ ), with those studies that did not report examining implementation having higher mean effect sizes (Mean ES = 0.20, 95% CI 0.05–0.35,  $k = 11$ ) than those studies that did (Mean ES = 0.05, 95% CI 0.02 to 0.10,  $k = 12$ ). This difference, however, failed to be sustained when controlling for target problem addressed ( $p = .96$ ) and level of the intervention ( $p = .30$ ), as lack of use of measures to assess for consistent implementation was correlated with an internalizing or broad socio-emotional focus ( $r = 0.70$ ) and with universal rather than selected implementation ( $r = 0.15$ ).

Of the 12 samples with calculated effect sizes that examined consistent program implementation, four reported implementation as adequate, seven reported high implementation quality, and one reported low implementation quality. There was no significant difference between those reporting adequate and those reporting high implementation quality ( $p = .75$ ).

**Inclusion of Parents, Teachers, and Peers.** The majority of programs included parents, teachers, or peers (78%). There were significant differences among studies that were ecologically guided and those that were not ( $Q_b = 5.86$ ,  $df = 1$ ,  $p = .016$ ). Those that did not target the ecology of the child had higher mean effect sizes (Mean ES = 0.36, 95% CI 0.11–0.61,  $k = 5$ ) than those that did reach beyond delivery at the individual level (Mean ES = 0.02, 95% CI 0.10 to 0.13,  $k = 18$ ). This difference no longer reached or approached significance, however, when controlling for target problem addressed ( $p = .90$ ) or intervention level ( $p = .13$ ), owing to a trend for programs with a conduct problems or substance use focus and those that were selective (each of which had weaker effects) to also be ecologically guided ( $r_s = 0.62$  and  $0.22$ , respectively).

**Use of Multiple Program Components.** The majority of programs included program components outside of the school-based curriculum (65%). There was no significant difference between effect sizes for evaluations of programs that incorporated components beyond those that were school based and those that did not ( $Q_b = 1.67$ ,  $df = 1$ ,  $p = .28$ ).

**Integration of Program Content Into General Classroom Curriculum.** Over half of the programs were delivered to samples in the classroom setting (56%). No significant difference in effect size was found as a function of this variable ( $Q_b = 1.00$ ,  $df = 1$ ,  $p = .32$ ).

**Developmental Timing.** The majority of samples were made up mostly of elementary students ( $k = 11$ ; 48%) or middle school students ( $k = 8$ ; 35%). Only four samples (17%) were high school students. Results revealed no main effect for age as a continuous ( $Q_b = 0.03$ ,  $df = 1$ ,  $p = .86$ ) or categorical variable (elementary vs. middle vs. high school;  $Q_b = 0.06$ ,  $df = 2$ ,  $p = .97$ ).

**Supplementary Analyses.** In view of the independent contributions of both target problem addressed and level of intervention to prediction of effect size, it was of interest to examine average effects for programs that reflected different combinations of these two program characteristics. As expected, there was a significant

difference among the four possible groups ( $Q_b = 39.21$ ,  $df = 3$ ,  $p < .001$ ), and the strongest effects were evident for programs that had an internalizing or broad/socio-emotional focus and were at a universal level (Mean ES = 0.33, 95% CI 0.22–0.45,  $k = 7$ ). Programs that had an internalizing or broad/socio-emotional focus and were at a selected/indicated level (Mean ES = 0.20, 95% CI 0.03 to 0.43,  $k = 4$ ) and those that had an externalizing or substance use focus and were at the universal level (Mean ES = 0.06, 95% CI 0.07 to 0.19,  $k = 3$ ) had positive but nonsignificant effects, whereas the remaining programs that had an externalizing or substance use focus and were at a selected/indicated level had a negative and significant effect (Mean ES = -0.14, 95% CI -0.23 to -0.04,  $k = 9$ ). The interaction between the two program characteristics in predicting effect size was nonsignificant ( $p = .66$ ). Of note also is that the prediction model taking into account both program characteristics was found to account for nearly all of the variance around the true population effect size (89.4%; Raudenbush, 1994). The residual within-group variation, furthermore, was nonsignificant for this model ( $Q_w = 24.82$ ,  $df = 19$ ,  $p = .17$ ). These findings suggest that after accounting for these two characteristics of programs, there was relatively little remaining variation in effect sizes.

## DISCUSSION

The goal of this review was to evaluate the effectiveness of school-based programs published in a 25-year period between 1985 and 2009 and targeting socio-emotional and behavioral outcomes among low-income, urban children and adolescents and to compare the results of this evaluation with results of an evaluation similar in scope but focused on the broader child and adolescent population (Rones & Hoagwood, 2000). Results of the current review reveal different patterns for low-income, urban youth in comparison with those reported for the broader population of youth, illustrating the importance of considering contextual influences on treatment effectiveness.<sup>3</sup>

The most basic difference between the results of the Rones and Hoagwood (2000) review, which included youth from all racial/ethnic, geographic, and socioeconomic backgrounds, and the results of the current review focused on low-income, urban youth is that

Rones and Hoagwood found greater evidence for effective school-based interventions. A direct comparison of results using the same qualitative methods revealed that Rones and Hoagwood (2000) found more programs to be effective and fewer to be ineffective than we found in our review. Results of our meta-analysis are consistent with this pattern, as the overall effect size for programs implemented with low-income, urban samples was very small (i.e., only 0.08 at post-treatment). Although a meta-analysis comparable in scope has not been conducted with the broader population of youth, these effects fall well below the medium to large effect sizes that have been reported for school-based prevention and intervention programs targeting youth from all backgrounds with established mental health problems (Reddy, Newman, Thomas, & Chun, 2009) and for psychotherapy treatments administered to the broader youth population (Weisz, Weiss, Han, Granger, & Morton, 1995). The results of this review highlight the need to more systematically evaluate the impact of socioeconomic factors on program development, mode of delivery, and treatment efficacy. Unfortunately, it has been reported that over 70% of all treatment outcome studies do not report on the socioeconomic characteristics of the participants (Weisz, Doss, & Hawley, 2005).

Of the 23 samples examined, only six provided follow-up data, and the average effect size at this time point was slightly smaller than at post. Only two studies reported effects at a second follow-up time point; therefore, interpretation of findings is limited. Nevertheless, analyses of differences for the same measures across time points revealed maintenance of gains over time for the intervention group in comparison with the control.

In an effort to understand what factors might be contributing to reduced program effectiveness for low-income, urban youth, we examined program target, level of intervention, provider of services, and race/ethnicity as possible moderators in our meta-analysis. With regard to target, results of moderation analyses revealed a significant difference in effect size between programs focused on externalizing problems (substance use and conduct problems) versus those focused on internalizing problems (depression and broad mental health and behavioral targets), with externalizing focused programs having a negative differential effect between treatment and

control groups and internalizing focused programs having much stronger and close to medium-sized differential treatment effects. Comparison of effect sizes for secondary outcomes revealed a similar pattern, with negative effects found for externalizing outcomes and positive effects found for internalizing/competence outcomes and academic outcomes.<sup>4</sup> Results of qualitative analyses also revealed similar effects. No programs specifically focused on conduct problems were deemed effective, 25% reported mixed results, and 75% were not effective. In contrast, samples included in Rones and Hoagwood's (2000) review (excluding low-income, urban samples) found 42% of conduct-focused programs to be effective and 21% not effective, with the remaining reporting mixed effects.

One interpretation of this pattern is that stressors endemic to urban poverty, such as exposure to community violence (Cooley-Quille, Boyd, Frantz, & Walsh, 2001), pull for externalizing problems, in particular (Grant et al., 2004), and interact with treatment to limit effectiveness for this outcome. A number of studies have reported high rates of externalizing problems relative to internalizing problems within the context of urban poverty (Grant et al., 2004). Furthermore, some emerging evidence suggests that exposure to community violence more strongly predicts externalizing relative to internalizing problems within this context (Grant et al., 2011). Scholars have theorized that externalizing problems may provide benefits within the context of urban poverty that include self-esteem, power, prestige, and protection from victimization (Cassidy & Stevenson, 2005). Additional research is needed to test the extent to which potential benefits associated with externalizing outcomes limit the effectiveness of interventions focused on this outcome for low-income, urban youth.

The prevalence of externalizing problems in low-income, urban settings may be the reason for the paucity of programs (which met inclusion criteria for our review) that focused on internalizing symptoms. Only two programs focused on depression, and no programs targeted other internalizing problems such as generalized anxiety or symptoms of post-traumatic stress disorder (PTSD).<sup>5</sup> Instead, most programs focused on conduct problems, substance use, or emotional/behavioral difficulties more generally, although one did

examine a treatment for concurrent internalizing and externalizing problems (Weiss, Harris, Catron, & Han, 2003). Given the established risk for trauma exposure among youth in urban centers of the United States (Kliewer et al., 2006), it is problematic that programs addressing PTSD and other internalizing symptoms among low-income, urban youth have not been systematically evaluated. Furthermore, the results of the current review suggest that these types of outcomes may be relatively amenable to treatment.

Results of moderation analyses with level of intervention as the moderator revealed an overall significant difference, with universal programs having positive effects and targeted programs having negative effects. Furthermore, supplemental analyses revealed that programs with an externalizing or substance use focus which were at a targeted/indicated level had significantly more negative effects than those administered at a universal level (which had positive but nonsignificant effects). This pattern is consistent with prior research showing that conduct-disordered youth may escalate their problem behavior in the context of interventions delivered in groups with similar peers (e.g., Dishion, McCord, & Poulin, 1999). At least two explanations have been provided for such "deviancy training" in the literature. The first is that youths' deviant behavior may be actively reinforced by externalizing peers' laughter, social attention, and interest. Another explanation is that youth may derive meaning and value from the material discussed in the program, which may pique interest or provide motivation to commit delinquent acts in the future (e.g., Dishion et al., 1999).

Prior reviews focusing on depression prevention programs have generally found that targeted programs are more effective than universal programs (Horowitz & Garber, 2006). However, universal school-based programs focused on youth anxiety have been found to be as effective or more effective than their targeted counterparts (Neil & Christensen, 2009). Reviews of school-based prevention programs for externalizing problems have similarly yielded inconsistent results regarding this issue. A recent meta-analysis of school-based randomized controlled trials focused on violence prevention did not find significant effects for either universal or targeted programs (Park-Higgerson,

Perumean-Chaney, Bartolucci, Grimley, & Singh, 2008). Wilson and Lipsey (2007), on the other hand, synthesized the findings of 249 experimental and quasi-experimental trials of school-based programs for aggressive and disruptive behavior and found favorable significant positive effects for both universal and targeted interventions. Consistent with the findings in this meta-analysis, Wilson and Lipsey also found that, among universal programs, those conducted with low-income samples had stronger effects, relative to those with mixed-income or middle-class samples. Income was not associated with differential effects when examined in the selected/targeted programs.

Results of moderation analyses with change agent as the moderator failed to reveal an overall significant difference as a function of who administered the intervention once target problem and level of the intervention were controlled. This is not consistent with some prior research. In particular, Payton et al. (2008), in their meta-analysis of social-emotional learning programs, found stronger overall effects for teacher-implemented programs than for researcher-implemented programs among a more general sample of youth. The contrasting findings for the current sample of low-income, urban youth may indicate another contextually specific influence. Urban poverty not only affects individuals directly but also influences them by compromising systems located in low-income communities (Conger et al., 1994). For example, schools in low-income, urban neighborhoods are often under-funded, under-resourced, and poorly functioning (e.g., Anyon, 1995). These systemic stressors, in turn, affect those individuals working within the school system (e.g., leading to high teacher stress, burnout, and turnover; e.g., Abel & Sewell, 1999). Such processes might limit the ability of individuals working within those systems to effectively implement interventions.

Results of moderation analyses with sample race/ethnicity as the moderator revealed no significant difference as a function of race/ethnicity. It is important to note, however, that low sample size for specific groups of youth (e.g., predominantly Latino youth) kept us from fully testing race/ethnicity as a moderator. In fact, as a result of sample size constraints, we were forced to combine two groups with substantially different effect sizes (i.e., predominantly African American

and Latino youth). Our resulting finding, therefore, that effect sizes for programs in which participating youth were predominantly of a single race/ethnicity did not differ from effect sizes for programs in which participating youth were of multiple racial/ethnic backgrounds does not provide us with information about possible differences in effects as a function of membership in specific racial/ethnic groups. Results of our analyses do indicate no evidence that including youth of predominantly one race/ethnicity, in general, is associated with intervention outcomes.

An examination of the five factors identified by Rones and Hoagwood (2000) as affecting outcomes, service sustainability, and maintenance revealed the following: (a) approximately half of the studies reported on implementation quality and, of these, no differential treatment effects emerged related to studies with high versus adequate implementation quality; (b) the majority of the programs in the current review included parents, teachers, and/or peers (78%), but this variable failed to function as a significant moderator once target problem and level of intervention were controlled; (c) a large portion (65%) of interventions included multiple components that reached beyond the school, but there were no significant differences between programs that did and did not do this; (d) over half of the programs were delivered in the classroom setting (56%), but there were no differences as a function of this variable; and (e) there were no differential effects based on age. A brief discussion of these findings follows.

In contrast with findings reported by Rones and Hoagwood (2000), significant differences between programs as a function of implementation quality did not emerge, and it is not clear why this was the case. Two possible explanations are provided. First, it may be that our measure of implementation, which was dependent upon investigators' reports of implementation assessment, was not valid. In other words, some investigators might have rigorously ensured fidelity but not reported it in their publication. An alternative explanation is that rigorous implementation of prescribed treatments may not always be warranted within the context of urban poverty. Perhaps flexible delivery and modification of treatment in real time bring benefits that temper the potential benefits of meticulous

implementation. Additional research is needed to test these hypothesized interpretations.

Also not consistent with Rones and Hoagwood's (2000) results were three findings related to ecological aspects of the interventions. Neither inclusion of parents, teachers, peers, nor use of multiple program components nor integration of program content into the classroom was found to influence program effectiveness. Although inconsistent with Rones and Hoagwood's results with majority youth, these findings are consistent with one another. They also are consistent with the finding discussed earlier that (in contrast with previously reported findings with majority samples; Payton et al., 2008) implementation by indigenous providers (e.g., teachers) was not associated with larger effect sizes relative to implementation by research staff. Taken together, these findings fit with the notion (also described earlier) that urban poverty not only affects youth directly but also influences them by compromising systems that affect them, such as families and schools (e.g., Abel & Sewell, 1999; Anyon, 1995; Gutman, McLoyd, & Tokoyawa, 2005). As mentioned, such processes may limit the ability of family members and school staff working within those systems to effectively implement interventions.

Finally, insufficient information was available to adequately assess the extent to which program components were developmentally appropriate; therefore, this variable was not analyzed in the same way it was in the Rones and Hoagwood (2000) review. Analyses of age as a moderator, however, revealed no significant effects, suggesting that developmental differences did not drive treatment effects in this sample of studies.

#### **Limitations**

One limitation of the current review is that our commitment to replicating the inclusion criteria used by Rones and Hoagwood (2000) meant that we included only published studies. This may have resulted in a bias toward significant findings (Reed, 2009). A second limitation that might also contribute to Type 1 error and affects all reviews is that authors might not have reported on all outcomes examined. In particular, authors may have been more likely to report positive and significant findings. Taken together, these limitations suggest that the current review's estimates of the percentage of effective

programs and of the effect sizes for these programs may be inaccurately high. Given that relatively few programs were found to be effective and the overall mean effect size for programs reviewed was very low, with confidence intervals indicating the effect may actually fall in the negative range, this limitation underscores the need for additional intervention development/modification work with low-income, urban youth.

Additionally, it is important to note that, despite the reported growth of SBHCs, the focus of the evaluations included in this review were not of services provided at these centers, but were instead of programs developed by researchers and delivered in school settings. One, however, included an evaluation of a treatment intervention for depression (Interpersonal Psychotherapy for Adolescents; IPT-A), as delivered by staff at an existing school-based mental health center versus treatment as usual (Mufson et al., 2004), and one was an evaluation of an intervention that was developed as part of the process of establishing a series of school-based mental health clinics (Weiss et al., 2003). The lack of evaluations of existing school-based mental health service delivery models suggests that rigorous research studies are simply not being conducted on mental health services delivered within school settings.<sup>6</sup> This represents an important gap in the research literature.

#### **IMPLICATIONS FOR RESEARCH AND PRACTICE**

Results of the current review raise questions about the validity of current models of change with underserved populations and highlight the need for additional intervention development work for youth living in urban poverty. The overall extremely small mean effect size across studies, the limited number of positive findings when directly compared with the Rones and Hoagwood (2000) results using the same qualitative approach, and the disconcerting presence of potential negative effects indicate that we remain very far from providing effective school-based services to low-income, urban youth.

Moderation findings provide leads for where to begin with this work. For example, basic and intervention development research should examine contextual processes that may pull for externalizing outcomes, in particular, and interact with treatment variables to limit positive effects for this outcome. Such work might

include the use of process observations, interviews, and focus groups, as well as quantitative analyses in order to develop innovative logic models for intervention that take influential contextual processes into account.

Until additional school-based interventions are developed, modified, and/or implemented effectively with this population, results of the current review suggest that caution should be used when implementing school-based interventions with low-income, urban youth that target conduct disorder or substance use symptoms, especially among youth already displaying externalizing problems. In addition, clinicians can refer to Table 1 for a small list of programs that were generally found to have larger effect sizes, and a larger list of programs that reported positive effects for some specific outcomes. The table also provides a list of interventions to avoid because of the potential for negative effects.<sup>1</sup>

In conclusion, the current review found limited evidence of effective school-based interventions for low-income, urban youth, especially for externalizing outcomes. Nonetheless, school-based services relative to other mental health service delivery programs, particularly for low-income, urban youth, have been shown to be more accessible and engaging (Atkins et al., 2006). This, coupled with promising results for a handful of interventions, highlights the potential of school-based mental health programs to meet the needs of youth living in urban poverty.

#### NOTES

1. A table outlining results of studies for which effect sizes could not be calculated is available from the first author.

2. Other controlled evaluations of school-based programs administered to low-income, urban youth (e.g., Dilworth, Mokrie, & Elias, 2002; Heydenberk & Heydenberk, 2005; Mayers, Hager-Bundy, & Buckner, 2008; Roberts, White, & Yeomans, 2004; Weisz et al., 2001) were excluded for lack of pre-post evaluation methods or lack of standardized outcomes, or for design issues (not randomized, quasi-experimental, or multiple baseline).

3. Weisz et al. (2005) summarized the youth treatment outcome literature across settings and conditions and found that very few reported sample demographic characteristics. Future reviews rely on these data in order to evaluate the performance of these interventions with youth of diverse backgrounds, including those with few economic resources in particular.

4. Program effects for primary, targeted outcomes were smaller than for secondary outcomes (although even these remained small at post-treatment and deteriorated further over time). Perhaps some youth in group interventions undermine explicit directions provided by the intervention in order to elicit attention from peers (Dishion et al., 1999) but unwittingly benefit from other aspects of the intervention that affect outcomes not overtly identified.

5. Although there have been published studies on PTSD among urban youth, primarily by the Cognitive Behavioral Intervention for Trauma in Schools (CBITS) program (i.e., Kataoka et al., 2003), studies were excluded when samples did not meet the low-income inclusion criteria.

6. Other published evaluations of these services with this population (e.g., Robinson, Harper, & Schoeny, 2003) were not included because of lack of pre-post designs.

#### ACKNOWLEDGMENT

This research was supported in part by a grant from the Institute of Education Sciences, the research arm of the United States Department of Education (Award Number: R305A080562).

#### REFERENCES

- \*Studies that met eligibility criteria for this review.
- Abel, M. H., & Sewell, J. (1999). Stress and burnout in rural and urban secondary school teachers. *Journal of Educational Research*, 92, 287-293.
- \*Aber, J. L., Jones, S. M., Brown, J. L., Chaudry, N., & Samples, F. (1998). Resolving conflict creatively: Evaluating the developmental effects of a school-based violence prevention program in neighborhood and classroom contexts. *Development and Psychopathology*, 10, 187-213.
- \*Abt Associates. (2001). *National evaluation of family support programs: Volume B. Research studies: Final report*. Cambridge, MA. Retrieved from <http://www.abtassociates.com/reports/NEFSP-VolB.pdf>
- Achenbach, T. M. (1991). *Manual for the Youth Self-Report and 1991 Profile*. Burlington: University of Vermont Department of Psychiatry.
- Anyon, J. (1995). Race, social class, and educational reform in an inner-city school. *Teacher College Record*, 97, 69-94.
- Atkins, M., Frazier, S., Birman, D., Adil, J. A., Jackson, M., Graczyk, P., et al. (2006). School-based mental health services for children living in high poverty urban communities. *Administration and Policy in Mental Health and Mental Health Services Research*, 33, 146-159.



- Bell, C. C., & Jenkins, E. (1994). Effects of child abuse and race. *Journal of the National Medical Association*, 86, 165-232.
- \*Botvin, G. J., Griffin, K. W., Diaz, T., & Iffill-Williams, M. (2001). Drug abuse prevention among minority adolescents: Posttest and one-year follow-up of school-based preventive intervention. *Prevention Science*, 2, 1-13.
- \*Cardemil, E. V., Reivich, K. J., Beevers, C. G., Seligman, M. E. P., & James, J. (2007). The prevention of depressive symptoms in low-income, minority children: Two-year follow-up. *Behaviour Research and Therapy*, 45, 313-327.
- \*Cardemil, E. V., Reivich, K. J., & Seligman, M. E. P. (2002). The prevention of depressive symptoms in low-income minority middle school students. *Prevention and Treatment*, 5: Article 8. Retrieved from <http://journals.apa.org/prevention/volume5/pre0050008a.html>.
- Cassidy, E. F., & Stevenson, H. C. J. (2005). They wear the mask: Hypervulnerability and hypermasculine aggression among African American males in an urban remedial disciplinary school. *Journal of Aggression, Maltreatment, and Trauma*, 11, 53-74.
- \*Cho, H., Hallfors, D. D., & Sanchez, V. (2005). Evaluation of a high school peer group intervention for at-risk youth. *Journal of Abnormal Child Psychology*, 33, 363-374.
- Conger, R. D., Ge, X., Elder, G. H., Jr., Lorenz, F. O., & Simons, R. L. (1994). Economic stressors, coercive family process, and developmental problems of adolescents. *Child Development*, 65, 541-561.
- Cooley-Quille, M., Boyd, R. C., Frantz, E., & Walsh, J. (2001). Emotional and behavioral impact of exposure to community violence in inner-city adolescents. *Journal of Clinical Child Psychology*, 30, 199-206.
- Cooper, H. (2010). *Research synthesis and meta-analysis: A step-by-step approach* (4th ed.). Thousand Oaks, CA: Sage.
- Dilworth, J., Mokruue, K., & Elias, M. (2002). The efficacy of a video-based teamwork-building series with urban elementary school students: A pilot investigation. *Journal of School Psychology*, 40, 329-346.
- Dishion, T. J., McCord, J., & Poulin, F. (1999). When interventions harm: Peer groups and problem behavior. *American Psychologist*, 54, 755-764.
- Dolan, L. J. (1992). *Models for integrating human services into the school* (Technical Report No. 30). Baltimore: Johns Hopkins University, Center for Research on Effective Schooling for Disadvantaged Students.
- Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41, 327-350.
- Farmer, E. M. Z., Burns, B. J., Phillips, S. D., Angold, A., & Costello, E. J. (2003). Pathways into and through mental health services for children and adolescents. *Psychiatric Services*, 54, 60-66.
- Farmer, E. M. Z., Stangl, D. K., Burns, B. J., Costello, E. J., & Angold, A. (1999). Use, persistence, and intensity: Patterns of care for children's mental health across one year. *Community Mental Health Journal*, 35, 31-46.
- Garland, A. F., Lau, A. S., Yeh, M., McCabe, K., Hough, R. L., & Landsverk, J. (2005). Racial and ethnic differences in utilization of mental health services among high-risk youths. *American Journal of Psychiatry*, 162, 1336-1343.
- Grant, K. E., Compas, B. E., Stuhlmacher, A., Thurm, A. E., McMahon, S., & Halpert, J. (2003). Stressors and child and adolescent psychopathology: Moving from markers to mechanisms of risk. *Psychological Bulletin*, 129, 447-466.
- Grant, K. E., Katz, B. N., Thomas, K. J., O'Koon, J. H., Meza, C. M., DiPasquale, A. M., et al. (2004). Psychological symptoms affecting low-income urban youth. *Journal of Adolescent Research*, 19, 613-634.
- Grant, K. E., Thomas, K. J., Carlson, G. C., Gipson, P. Y., Mance, G. A., Carter, J. S., et al. (2011). *Specific processes linking economic stressors, exposure to community violence, and psychological symptoms affecting low-income urban youth*. Manuscript submitted for publication.
- Gutman, L., McLoyd, V. C., & Tokoyawa, T. (2005). Financial strain, neighborhood stress, parenting behaviors, and adolescent adjustment in urban African American families. *Journal of Research on Adolescence*, 15, 425-449.
- Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6, 107-128.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. New York: Academic Press.
- Hedges, L. V., & Vevea, J. L. (1998). Fixed and random-effects models in meta-analysis. *Psychological Methods*, 3, 486-504.
- \*Henderson, P., Kelbey, T., & Engerbreston, K. (1992). Effects of a stress-control program on children's locus of control, self-concept, and coping behavior. *The School Counselor*, 40, 125-130.
- Heydenberk, B. A., & Heydenberk, W. R. (2005). Increasing meta-cognitive competence through conflict resolution. *Education and Urban Society*, 37, 431-445.
- Higgins, J. P. T., & Green, S. (2008). *Cochrane handbook for systematic reviews of interventions*. The Cochrane collaboration. Retrieved from <http://www.cochrane-handbook.org/>
- Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistical Methods*, 21, 1539-1558.

- Horowitz, J. L., & Garber, J. (2006). The prevention of depressive symptoms in children and adolescents: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 74*, 401–415.
- \*Hostetler, M., & Fisher, K. (1997). Project C.A.R.E. substance abuse prevention program for high-risk youth: A longitudinal evaluation of program effectiveness. *Journal of Community Psychology, 25*, 397–419.
- Institute of Medicine. (1994). *Reducing risks for mental disorders: Frontiers for preventive intervention research*. Washington, DC: National Academy Press.
- \*Jagers, R. J., Morgan-Lopez, A. A., Howard, T. L., Browne, D. C., & Flay, B. R. (Aban Aya Coinvestigators). (2007). Mediators of the development and prevention of violent behavior. *Prevention Science, 8*, 171–179.
- Kataoka, S. H., Stein, B. D., Jaycox, L. H., Wong, M., Escudero, P., Tu, W., et al. (2003). A school-based mental health program for traumatized Latino immigrant children. *Journal of the American Academy of Child and Adolescent Psychiatry, 42*, 311–318.
- Kataoka, S. H., Zhang, L., & Wells, K. B. (2002). Unmet need for mental health care among US children: Variation by ethnicity and insurance status. *American Journal of Psychiatry, 159*, 1548–1555.
- Kazdin, A. E., Holland, L., & Crowley, M. (1997). Family experience of barriers to treatment and premature termination from child therapy. *Journal of Consulting and Clinical Psychology, 65*, 453–463.
- Kliewer, W., Parrish, K. A., Taylor, K. W., Jackson, K., Walker, J. M., & Shivy, V. A. (2006). Socialization of coping with community violence: Influences of caregiver coaching, modeling, and family context. *Child Development, 77*, 605–623.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
- Mayers, H. A., Hager-Bundy, M., & Buckner, E. B. (2008). The Chances for Children Teen Parent-Infant Project: Results of a pilot intervention for teen mothers and their infants in inner city high schools. *Infant Mental Health Journal, 29*, 320–342.
- \*McClowry, S., Snow, D. L., & Tannis-LeMonda, C. S. (2005). An evaluation of the effects of "INSIGHTS" on the behavior of inner city primary school children. *Journal of Primary Prevention, 26*, 567–583.
- \*McDonald, L., Moberg, D. P., Brown, R., Rodriguez-Espiricueta, I., Flores, N. I., Burke, M. P., et al. (2006). After-school multifamily groups: A randomized controlled trial involving low-income, urban, Latino children. *Children and Schools, 28*, 25–36.
- McHugh, R., Murray, H. W., & Barlow, D. H. (2009). Balancing fidelity and adaptation in the dissemination of empirically-supported treatments: The promise of transdiagnostic interventions. *Behaviour Research and Therapy, 47*, 946–953.
- McMiller, W. P., & Weisz, J. R. (1996). Help-seeking preceding mental health clinic intake among African-American Latino, and Caucasian youths. *Journal of the American Academy of Child & Adolescent Psychiatry, 35*, 1086–1094.
- \*Metropolitan Area Child Study Research Group. (2002). A cognitive-ecological approach to preventing aggression in urban settings: Initial outcomes for high-risk children. *Journal of Consulting and Clinical Psychology, 70*, 179–194.
- \*Mufson, L., Dorta, K. P., Wickramaratne, P., Nomura, Y., Olsson, M., & Weissman, M. M. (2004). A randomized effectiveness trial of interpersonal psychotherapy for depressed adolescents. *Archives of General Psychiatry, 61*, 577–584.
- \*Murray, C., & Malmgren, K. (2005). Implementing a teacher-student relationship program in a high-poverty urban school: Effects on social, emotional, and academic adjustment and lessons learned. *Journal of School Psychology, 43*, 137–152.
- National Assembly on School-Based Health Care. (2007). *School-based health centers national census school year 2004–05*. Retrieved from: <http://www.nasbhc.org/atf/cf/%7BCD9949F2-2761-42FB-B3C7A-CEE165C7011D9%71D/Census2005.pdf>
- Neil, A. L., & Christensen, H. (2009). Efficacy and effectiveness of school-based prevention and early intervention programs for anxiety. *Clinical Psychology Review, 29*, 208–215.
- Park-Higginson, H. K., Perumean-Chaney, S. E., Bartolucci, A. A., Grimley, D. M., & Singh, K. P. (2008). The evaluation of school-based violence prevention programs. A meta-analysis. *Journal of School Health, 78*(9), 465–479.
- Paternite, C. E. (2005). School-based mental health programs and services: Overview and introduction to the special issue. *Journal of Abnormal Child Psychology, 33*, 657–663.
- Payton, J., Weissberg, R. P., Durlak, J. A., Dymnicki, A. B., Taylor, R. D., Schellinger, K. B., et al. (2008). *The positive impact of social and emotional learning for kindergarten to eighth-grade students: Findings from three scientific reviews*. Chicago: Collaborative for Academic, Social, and Emotional Learning. Retrieved from <http://www.casel.org/downloads/PackardES.pdf>
- Raudenbush, S. W. (1994). Random effects models. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 301–321). New York: Russell Sage Foundation.

- Reddy, L. A., Newman, E., Thomas, C. A. D., & Chun, V. (2009). Effectiveness of school-based prevention and intervention programs for children and adolescents with emotional disturbances: A meta-analysis. *Journal of School Psychology, 47*, 77–99.
- Reed, J. (2009). A novel application of the file drawer formula: Rejected study bias in meta-analysis. *Psychological Reports, 105*, 703–706.
- Roberts, L., White, G., & Yeomans, P. (2004). Theory development and evaluation of project WIN: A violence reduction program for early adolescents. *Journal of Early Adolescence, 24*, 460–483.
- Robinson, W. L., Harper, G. W., & Schoeny, M. E. (2003). Reducing substance use among African-American adolescents: Effectiveness of school-based health centers. *Clinical Psychology: Science and Practice, 10*, 431–504.
- Rones, M., & Hoagwood, K. (2000). School-based mental health services: A research review. *Clinical Child and Family Psychology Review, 3*, 223–240.
- Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231–244). New York: Russell Sage Foundation.
- \*Sinclair, M. F., Christenson, S. L., & Thurlow, M. L. (2005). Promoting school completion of urban secondary youth with emotional or behavioral disabilities. *Exceptional Children, 71*, 465–482.
- \*Tolan, P., Gorman-Smith, D., & Henry, D. (2004). Supporting families in a high-risk setting: Proximal effects of the SAFEChildren prevention intervention. *Journal of Consulting and Clinical Psychology, 72*, 855–869.
- Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.
- U.S. Bureau of the Census. (2000). *Urbanized area and urban cluster information census*. Retrieved from <http://www.census.gov/geo/www/ua/uauinfo.html#lists>
- U.S. Bureau of the Census. (2011). *Income, expenditures, poverty, & wealth*. Washington, DC: Government Printing Office. Retrieved from [http://www.census.gov/compendia/statab/cats/income\\_expenditures\\_poverty\\_wealth.html](http://www.census.gov/compendia/statab/cats/income_expenditures_poverty_wealth.html)
- U.S. Department of Health and Human Services. (1999). Chapter 3: Children and mental health. In *Mental health: A report of the surgeon general*. Rockville, MD: U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Center for Mental Health Services.
- U.S. Department of Health and Human Services. (2001). Chapter 2: Culture counts: The influence of culture and society on mental health. In *Mental health: Culture, race, and ethnicity—A supplement to mental health: A report of the surgeon general*. Rockville, MD: U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Center for Mental Health Services.
- Walker, S. C., Kerns, S. E. U., Lyon, A. R., Bruns, E. J., & Cosgrove, T. J. (2010). Impact of school-based health center use on academic outcomes. *Journal of Adolescent Health Care, 46*, 251–257.
- \*Weiss, B., Harris, V., Catron, T., & Han, S. (2003). Efficacy of the RECAP intervention program for children with concurrent internalizing and externalizing problems. *Journal of Consulting and Clinical Psychology, 71*, 364–374.
- Weisz, J. R., Doss, A., & Hawley, K. M. (2005). Youth psychotherapy outcome research: A review and critique of the evidence base. *Annual Review of Psychology, 56*, 337–363.
- Weisz, J., Weiss, B., Han, S., Granger, D., & Morton, T. (1995). Effects of psychotherapy with children and adolescents revisited: A meta-analysis of treatment outcome studies. *Psychological Bulletin, 117*, 450–468.
- What Works Clearinghouse. (2008). *Procedures and standards; Appendix C—Clustering correction of the statistical significance of effects estimated with mismatched analyses*. Retrieved from <http://ies.ed.gov/ncee/wwc/references/iDocViewer/Doc.aspx?docId=19&tocId=9>
- Wheeler, M. E., Keller, T. E., & DuBois, D. L. (2010). Review of three recent randomized trials of school-based mentoring. *Social Policy Report, 24*, 1–21.
- Wilson, S. J., & Lipsey, M. W. (2007). School-based interventions for aggressive and disruptive behavior. Update of a meta-analysis. *American Journal of Preventive Medicine, 33*(2S), S130–S143.

Received September 19, 2011; accepted September 21, 2011.